



花椒窄吉丁转录组及化学感受相关基因的分析

杨 平, 谢寿安*, 巩雪芳, 车显荣, 王延来, 吕淑杰

(西北农林科技大学林学院, 陕西杨陵 712100)

摘要:【目的】建立花椒窄吉丁 *Agrilus zanthoxylumi* 转录组数据库, 挖掘其基因组数据。【方法】采用高通量测序平台 Illumina NovaSeq6000 对花椒窄吉丁成虫不同组织(触角、头、胸、腹、足、翅)进行转录组测序、序列组装, 使用 BLAST 软件将花椒窄吉丁 unigenes 序列与公共数据库进行比对, BLAST 同源性搜索的方法从中筛选出与其化学感受相关的基因, 并与其他已研究发表的鞘翅目昆虫化学感受相关基因的核酸序列比对, 进行系统发育分析。利用 RPKM 值对这些基因在花椒窄吉丁成虫不同组织中的表达量进行分析。【结果】经测序及序列拼接后共得到 80 320 条 unigenes 和 169 398 条 contigs, 其 G+C 比例分别是 37.63% 和 39.18%, 平均长度分别为 828.75 和 1 084.33 bp; unigenes 长度主要分布在 200~400 bp 的共有 37 374 条, 占全部 unigenes 的 46.53%。在 NR 数据库中注释的与花椒窄吉丁转录组相似基因所属物种分布为赤拟谷盗 *Tribolium castaneum* 所占比例 26.80%, 其后依次是小家鼠 *Mus musculus* (13.87%), 蛀犀金龟 *Oryctes borbonicus* (5.26%), 山松甲虫 *Dendroctonus ponderosae* (4.10%), 黑蚁 *Lasius niger* (2.36%), 其他物种所占比例是 47.61%。在 GO 数据库中比对到 27 488 个 unigenes, 可分为分子功能、细胞组分和生物学进程三大类共 59 个过程; 利用 KOG 数据库功能注释分为 25 类, 注释到 unigenes 最多功能类别的是普通功能, 共 4 666 个, 最少的是与细胞移动相关, 共 40 个; 将 80 320 个 unigenes 映射到 KEGG 数据库中, 21 104 条 unigenes 注释到代谢通路 35 个, 占 26.27%, 注释最多的代谢途径是转录途径, 共涉及到 2 037 条 unigenes。从 NR 数据库中注释到 8 个气味受体 (odorant receptor, OR) 基因、7 个离子型受体 (ionotropic receptor, IR) 基因和 6 个味觉受体 (gustatory receptor, GR) 基因, 各组织中的基因表达量分析发现, 这 3 类化学感受相关基因在成虫触角中的表达量显著高于其他组织中的表达量, 雌、雄触角中表达量差异最显著的为 *AzanOR1*, 而在雌、雄成虫足中这 3 类化学感受相关基因表达量没有差异。【结论】本研究首次获得了花椒窄吉丁转录组数据, 为进一步研究花椒窄吉丁的基因功能奠定了分子基础。

关键词: 花椒窄吉丁; 转录组; 高通量测序; 基因注释; 生物信息学

中图分类号: S763.38 **文献标识码:** A **文章编号:** 0454-6296(2019)05-0547-14

Analysis of the transcriptome and chemoreception-related genes of *Agrilus zanthoxylumi* (Coleoptera: Buprestidae)

YANG Ping, XIE Shou-An*, GONG Xue-Fang, CHE Xian-Rong, WANG Yan-Lai, LÜ Shu-Jie
(College of Forestry, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: 【Aim】To establish the transcriptome database of the *Agrilus zanthoxylumi*, and to obtain its genomic data. 【Methods】The transcriptome of the different tissues (antennae, head, thorax, abdomen, leg, wing) of *A. zanthoxylumi* adults was sequenced using an Illumina NovaSeq 6000 platform, and the unigene sequences were compared with the homologous sequences in public databases by BLAST.

Chemoreception-related genes were screened out by BLAST homology search, and their nucleotide sequences were compared with those of other reported chemoreception-related genes of Coleoptera insects for phylogenetic analysis. The expression levels of these genes in different tissues of *A. zanthoxylumi* adults were analyzed based on the RPKM value. 【Results】 After sequencing and sequence splicing, a total of 80 320 genes and 169 398 contigs were obtained. The G + C proportions of contigs and unigenes were 37.63% and 39.18%, respectively, and the average length of contigs and the average length of unigenes were 1 084.33 and 828.75 bp, respectively. There were 37 374 unigenes with a length distribution mainly in 200 – 400 bp, accounting for 46.53% of the total unigenes. Origin species distribution of similar genes with unigenes of *A. zanthoxylumi* annotated in the NR database was *Tribolium castaneum* (26.80%), *Mus musculus* (13.87%), *Oryctes borbonicus* (5.26%), *Dendroctonus ponderosae* (4.10%), *Lasius niger* (2.36%), and other species (47.61%). There were 27 488 unigenes aligned to the GO database, which can be divided into three functional categories including molecular function, cellular component and biological process in a total of 59 processes. According to the gene annotation against the KOG database, these unigenes can be divided into 25 categories. The most annotated functional category was the common function (4 666 unigenes), while the least annotated was cell movement-related (only 40 unigenes). When 80 320 unigenes were mapped to the KEGG database, 21 104 unigenes were grouped into 35 metabolic pathways, accounting for 26.27% of the total. The most annotated metabolic pathway was the transcription pathway (2 037 unigenes). By gene annotation against the NR database, we found eight odorant receptor (OR) genes, seven ionotropic receptor (IR) genes and six gustatory receptor (GR) genes. Tissue expression profiles showed that the three kinds of chemoreception-related genes (OR, IR and GR genes) had significantly higher expression levels in the antennae than in other tissues of adults, and the expression level of *AzanOR1* in the antennae showed the most significant difference between female and male, while the expression levels of the three kinds of chemoreception-related genes in the legs of adults showed no difference between female and male. 【Conclusion】 This study obtained the transcriptome data of *A. zanthoxylumi*. The results provide a molecular foundation for further studying the gene function in *A. zanthoxylumi*.

Key words: *Agilus zanthoxylumi*; transcriptome; HiSeq; gene annotation; bioinformatics

花椒窄吉丁 *Agilus zanthoxylumi* 隶属鞘翅目 (Coleoptera) 吉丁虫科 (Buprestidae) 窄吉丁属 *Agilus* (孙益知和李忠锋, 1988), 一化性昆虫, 均为一年发生 1 代, 主要分布在我国北方花椒产区, 比如陕西、甘肃、山西、山东等。目前研究发现花椒窄吉丁仅危害花椒, 为单食性昆虫 (吴海, 2006)。幼虫在木质部或树皮隧道中越冬, 成虫以叶片为食, 对花椒叶造成缺刻和空洞 (崔艳鱼, 1990), 受害花椒树皮的主要症状是会大量流胶、软化、腐烂等, 树皮干枯龟裂, 严重会导致花椒树体死亡 (何凯, 1991)。现阶段对花椒窄吉丁的研究主要集中于化学防治和生物学特性两方面 (康克功等, 2007), 利用天敌对其进行防控也有少量报道 (张伟, 2009), 而长期使用化学药剂防治的策略会使害虫产生 3R 问题 (张美翠等, 2014), 会导致运用化学防治也很难奏效, 还会影响花椒的产量及其商品价值, 给花椒产业造

成了重大的经济损失。生物防治作为应用较为广泛的无公害防治技术之一 (陈学新等, 2013)。随着我国绿色无公害生物防治技术的发展, 探究农林业有害昆虫的高效防控, 基因组信息不全已成为亟待解决的重要问题之一。然而, 关于花椒窄吉丁转录组信息方面的研究还是空白。

选用高通量转录组测序技术测定花椒窄吉丁转录组信息的原因, 主要是高通量转录组测序技术拥有精确到单个核苷酸的分辨率, 且能对样品任意时间点或任意条件下的转录组进行测序, 还能够动态反映基因转录水平 (周华等, 2012), 同时鉴定和定量稀有转录本和正常转录本, 提供样品特异的转录本序列结构信息和确定基因序列和转录本, 在没有参考基因的条件下也可以对某一物种的基因组进行细致全面的分析, 为基因组图谱尚未完成或数据信息匮乏的物种的基因组和转录组研究提供技术支持

等优势而在昆虫转录组学研究上得到广泛使用(贺华良等, 2012)。从 2008 年至今, 研究人员已采用新一代测序技术对 7 个目的 68 种昆虫进行了转录组测序, 其中由我国研究人员完成 6 个目的 22 种昆虫的转录组测序(杨帆等, 2014)。目前, 昆虫转录组学研究生物体的细胞或组织在特定状态下基因组所转录的全部 mRNA, 其反映了基因在不同生命阶段、生理状态、组织类型以及环境条件下的表达情况(Yang *et al.*, 2014)。对这些基因进行深入挖掘、基因表达分析后, 可为全面揭示昆虫生命活动中相关基因功能、进化以及昆虫与其他生物相互作用等奠定基础。

本研究采用 Illumina NovaSeq6000 测序平台对花椒窄吉丁成虫转录组测序、拼接组装得到 unigene。通过 unigene 与数据库比对, 获得基因功能注释、分类和代谢途径分析, 旨在获得更多鞘翅目非模式生物花椒窄吉丁的转录本和更为全面的转录组信息, 为生物控制花椒窄吉丁提供重要的理论依据和相关参考。于新一代测序技术的昆虫转录组学研究, 为昆虫分子标记的开发提供了丰富的遗传学资源(Gayral *et al.*, 2011)。袁丽芳(2016)利用固相微萃取和 GC-MS 分析了花椒健康植株和受花椒窄吉丁危害植株不同部位的主要挥发性化合物, 采用触角电位(EAG)和 Y 型嗅觉仪测定了花椒窄吉丁对寄主植物挥发物的电生理和行为反应。但花椒窄吉丁是如何感知寄主气味物质并准确定位寄主, 在寻找寄主植物过程中化学感受相关基因是如何发挥作用的, 目前还没有相关报道。为此, 本研究通过转录组数据中基因功能注释及分析挖掘到花椒窄吉丁的大量化学感受相关基因, 为进一步从分子水平开展花椒窄吉丁防治新策略的研究奠定理论基础。

1 材料与方法

1.1 供试昆虫

花椒窄吉丁成虫采集于陕西省西安市蓝田县普化镇周董村花椒园, 采集时间为 2018 年 5–7 月。借助捕虫工具寻找在叶片、枝干上栖息的成虫迅速将其逮住, 放入扎好透气孔的离心管中, 每个离心管塞入一小片花椒叶片。带回实验室置于 25℃ 下饲养, 选取 1 000 头活动能力较强的成虫进行虫体清洗之后在解剖镜下进行雌雄鉴别, 于冰上迅速进行各组织分离、收集, 迅速用液氮冷冻并保存于 -80℃ 低温冰箱中备用。

1.2 总 RNA 的提取与质量检测

采用 Trizol 法分别提取花椒窄吉丁成虫不同组织(触角、头、胸、腹、足、翅)的总 RNA, 每个样品设置 3 个生物学重复, 1% 琼脂糖凝胶电泳检测总 RNA 的质量。凯奥 K5500 分光光度计(凯奥, 北京); 检测 RNA 样品的纯度(OD_{260}/OD_{280} 值), 根据安捷伦 2100 RNA Nano 6000 Assay Kit (Agilent Technologies, CA, 美国)检测 RNA 样品的完整性和浓度。

1.3 文库构建与测序

1.2 节提取的总 RNA 样本检测合格后, 用带有 Oligo(dT) 的磁珠富集 mRNA, 向得到的 mRNA 中加入片段缓冲液使其片段化成为短片段, 再以片段化后的 mRNA 为模板, 用 6 碱基随机引物合成 cDNA 第 1 链, 并加入缓冲液、dNTPs、RNaseH 和 DNA Polymerase I 继续合成 cDNA 第 2 链, 经过 QIAQuickPCR 试剂盒纯化并加 EB 缓冲液洗脱。洗脱纯化后的双链 cDNA 再进行末端修复、加碱基 A、加测序接头处理, 然后经琼脂糖凝胶电泳回收目的大小片段并进行 PCR 扩增, 从而完成整个文库制备工作。构建好的文库用 Illumina NovaSeq6000 平台进行测序, 测序读长为 PE150。

1.4 数据过滤与拼接预组装

测序完成后获得原始 reads, 去除带接头、poly-N 和低质量的 reads, 获得 clean reads。每个碱基测序错误率是由测序 Phred 分值(Phred score, Qphred)通过公式转化得到, 而 Phred 分值是在碱基识别(base calling)过程中, 通过一种预测碱基判别发生错误概率模型计算 Q20, Q30 和 GC 含量和重复序列水平。采用 Trinity 软件对 clean reads 进行拼接、过滤和组装后获得高质量的 unigene。由于该样品昆虫花椒窄吉丁为无参考基因组序列的样品, 因此在获得 clean reads 后需要进行拼接, 从而获取后续分析所需的参考序列(Grabherr *et al.*, 2011; Hansen *et al.*, 2012)。

Trinity 利用 deBruijn 图论的基本原理, 组装出全长转录本。基于对 Trinity 组装出来的转录本序列, 取每个基因中最长的转录本序列作为 unigene, 再对 unigene 长度分布进行统计。将统计到的不同长度的 unigene 分别与 NR, KEGG 和 COG 数据库进行序列比对和信息分析(期望值 $E < 1.00E^{-05}$)。在此基础上完成全部 unigene 序列的方向定位和同源序列信息汇总。将不同样品组装结果中多个可变剪接的转录本用 Chrysalis 聚类得到 unigene 库。通

过 unigene 在 Mapped Reads 上的分布,检验片段化的随机程度、基因数目的饱和情况。测序数据量可正向反映测序基因的数目,数据量越大得到的基因数目越多。通过 unigene 核酸序列与数据库的 BLAST 比对,可得到 unigene 的蛋白功能注释信息,比对中选择参数 E-value 不大于 10^{-5} 。

1.5 unigene 的功能注释

采用一些主要功能数据库, Swiss-Prot, NR, NT, PFAM, GO, KO, eggNOG, BLASTP, BLASTX 等对基因和蛋白质功能注释。

1.6 花椒窄吉丁化学感受相关基因系统发育树分析

利用 NCBI 在线软件进行花椒窄吉丁 unigene 的 BLAST 比对,将比对相似度较高的鞘翅目昆虫的化学感受相关基因序列下载下来,利用 MEGA 软件构建系统发育树,比较花椒窄吉丁候选气味受体 (odorant receptor, OR)、离子型受体 (ionotropic receptor, IR) 和味觉受体 (gustatory receptor, GR) 基因与鞘翅目其他昆虫相关基因的近源关系。

1.7 花椒窄吉丁化学感受相关基因的表达量分析

通过 RPKM (reads per kilobase per million mapped reads) 值来衡量基因在花椒窄吉丁成虫不同组织中的表达水平,RPKM 是利用 RNA-Seq 技术用来定量估计基因表达值的有效工具,能够消除基因长度和测序量差异对计算基因表达的影响,计算得到的基因表达量可直接用于比较不同样品间的基因表达差异,此文中旨在比较不同组织中的基因表

达量差异。其计算公式为:

公式中 R 为唯一比对到某个基因的 reads 数, N 为唯一比对到基因的总 reads 数, L 为此基因的长度。

$$RPKM = \frac{10^6 \times R}{NL/10^3}。$$

2 结果

2.1 花椒窄吉丁转录组的测序和组装

利用 Illumina NovaSeq6000 平台测序,组装得到 80 320 条 unigenes 和 169 398 条 contigs(表 1), G + C 比例分别是 37.63% 和 39.18%, 碱基总数分别是 183 683 635 和 66 565 475。contigs 和 unigenes 的平均长度分别为 1 084.33 和 828.75 bp(表 2), N50 长度分别为 2 108 和 1 438 bp。长度在 200 ~ 400 bp 的 unigenes 有 37 374 条, 占全部 unigenes 的 46.53%; 而 400 ~ 600 bp 的有 14 108 条占 17.56%; 其余长度的 unigenes 数量占比均小于 10%(图 1), 表明测序结果和组装效果很好, 满足进一步开展基因功能分析要求。

表 1 花椒窄吉丁转录组组装结果统计
Table 1 Summary of assembly for the transcriptome of *Agrilus zanthoxylumi*

	数量 Number	(G + C) %	碱基总数 Total number of bases
Contigs	169 398	39.18	66 565 475
Unigenes	80 320	37.63	183 683 635

表 2 花椒窄吉丁转录组序列长度统计
Table 2 Statistics of sequence length of the transcriptome of *Agrilus zanthoxylumi*

	N50 (bp)	N90 (bp)	最小长度 (bp) Minimum length	最大长度 (bp) Maximal length	数量 Number	平均长度 (bp) Mean length
Contigs	2 108	399	201	28 219	169 398	1 084.33
Unigenes	1 438	313	201	28 219	80 320	828.75

N50: 序列从大到小排列,当其累计长度刚刚超过全部序列总长度 50% 时,最后一个序列的大小即为 N50 的大小 The sequences are arranged from large to small, and when the accumulative length just exceeds 50% of the total length of all sequences, the size of the last sequence is N50; N90: 序列从大到小排列,当其累计长度刚刚超过全部序列总长度 90% 时,最后一个序列的大小即为 N90 的大小 The sequences are arranged from large to small, and when the cumulative length just exceeds 90% of the total length of all the sequences, the size of the last sequence is the size of N90.

Q30 值为 99.9%, 说明碱基识别率较高,测序质量高数据可靠性好,对应关系如表 3 所示。

为保证分析准确性,以 Q30 碱基百分比作为指标进行统计 clean data 的质量, Q30 碱基百分比越大说明测序错误率小于 0.1% 的碱基在总碱基中的比例越大。图 2 以所有花椒窄吉丁不同组织 (触角、头、胸、腹、足、翅) 为横坐标,以 Q30 百分比作为对应的纵坐标直观地反映了最终过滤后的数据可靠

性。所有组织碱基百分比都在 91% 以上,说明测序数据可靠可进行后续分析。

2.2 花椒窄吉丁转录组基因功能注释

通过选择 BLAST 参数 $E\text{-value} \leq 1e^{-5}$ 和 HMMER 参数 $E\text{-value} \leq 1e^{-10}$, 最终获得了 80 320 个有注释信息的 unigenes, 数据库注释结果表明, NR 数据库注释 32 470 条 (40.43%), BLASTX 注释 29 382 条 (36.58%), GO 数据库注释 27 488 条

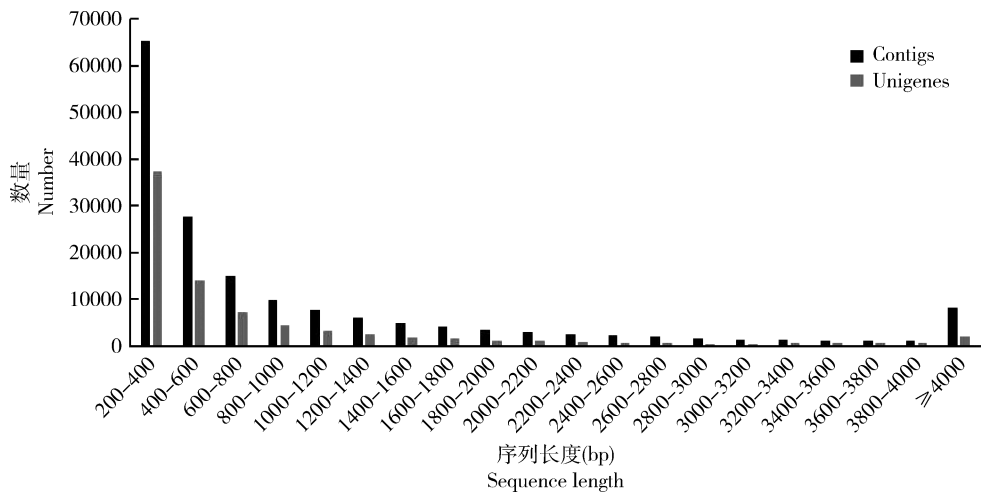


图 1 花椒窄吉丁转录组 contigs 和 unigenes 长度分布

Fig. 1 Length distribution of contigs and unigenes in the transcriptome of *Agrilus zanthoxylumi*

表 3 Illumina Casava 碱基识别与 Phred 分值之间的简明对应关系			
Table 3 The concise correspondence between Illumina Casava base recognition and Phred score			
Phred 分值	不正确的碱基识别	碱基正确识别率(%)	Q 分值
Phred score	Incorrect base-recognition	Correct base recognition rate	Q-score
10	1/10	90.00	Q10
20	1/100	99.00	Q20
30	1/1 000	99.90	Q30
40	1/10 000	99.99	Q40

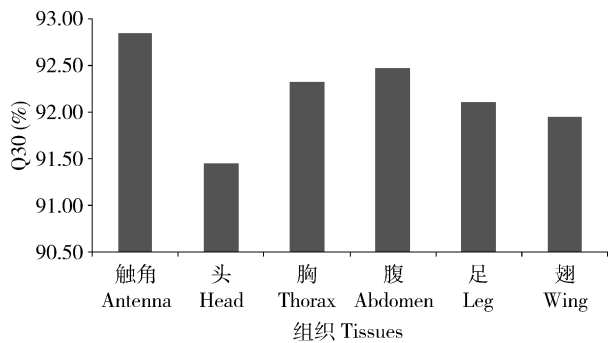


图 2 花椒窄吉丁转录组 clean data 的质量统计

Fig. 2 Quality statistics of clean data of the transcriptome of *Agrilus zanthoxylumi*

Q30: 过滤后,总序列中质量值大于 30 (错误率小于 0.1%) 的碱基数的比例 Percentage of bases with a quality score greater than or equal to 30 (incorrect base recognition rate: <0.1%).

(34.22%), SwissProt 库注释 25 941 条(32.30%), BLASTP 注释 18 921 条(23.56%), PFAM 数据库注释 18 038 条(22.46%), eggNOG 注释 18 013 条(22.43%), KO 数据库注释 17 312 条(21.55%)以及 NT 数据库注释 12 056 条(15.01%)(表 4)。

在 BLASTP, NR, NT 和 BLASTX 这 4 个数据库中 都得到注释的 unigene 数量是 6 221 条(图 3);在 NR, NT 和 BLASTX 这 3 个数据库注释的结果相对

表 4 花椒窄吉丁转录组基因功能注释结果统计		
Table 4 Statistics of unigene annotation of the transcriptome of <i>Agrilus zanthoxylumi</i>		
数据库	Unigene 数量	百分比
Database	Number of unigenes	Percentage
Swiss-Prot	25 941	32.30
NR	32 470	40.43
NT	12 056	15.01
PFAM	18 038	22.46
GO	27 488	34.22
KO	17 312	21.55
eggNOG	18 013	22.43
BLASTP	18 921	23.56
BLASTX	29 382	36.58

比较多,所测得到的花椒窄吉丁的 unigene 数量能在现阶段各数据库中都注释到一定的数量。

通过与 NR 数据库比对注释,花椒窄吉丁转录组注释到其他物种分布图显示(图 4),与赤拟谷盗 *Tribolium castaneum* 的相似基因所占比例较高(26.80%),其后依次是小家鼠 *Mus musculus* (13.87%), 蛀犀金龟 *Oryctes borbonicus* (5.26%), 山松甲虫 *Dendroctonus ponderosae* (4.10%), 黑蚁 *Lasius niger* (2.36%), 其他物种占 47.61%。

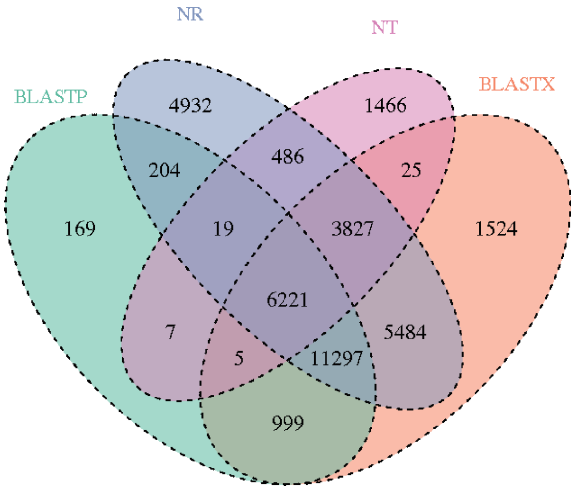


图3 花椒窄吉丁转录组基因功能注释结果韦恩图
Fig. 3 Venn diagram of the annotation results of the transcriptome of *Agrilus zanthoxylumi*

图中数值表示注释到某个数据库的花椒窄吉丁 unigene 数量,交集部分的数值则表示同时注释在重叠部分数据库的 unigene 数量。The numerals in the figure represent the number of unigenes annotated into a certain database, while the numerical value of the intersection part indicates the number of unigenes annotated in the overlapping part of the database.

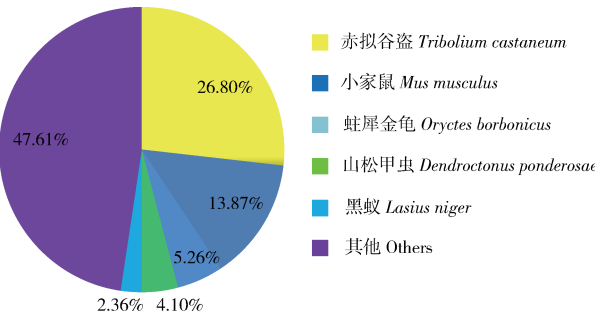


图4 NR 数据库中注释的花椒窄吉丁 unigenes 的物种分布
Fig. 4 Species distribution of *Agrilus zanthoxylumi* unigenes in the NR database

2.3 花椒窄吉丁 unigenes 的功能分类

经 GO 数据库比对到 27 488 个 unigenes 可分 3 个功能类别:分子功能 (molecular function)、细胞组分 (cellular component) 和生物学进程 (biological process) 三大类共 59 个过程 (图 5)。从图中结果可知参与生物学进程的 unigenes 有 27 类,其中参与细胞进程的 unigenes 占总体的 73.24%,而代谢进程的占 60.10%,这两个过程所占比例最多;其次是生物调控和细胞组分组织或生物合成分别占 38.22% 和 22.71%。细胞组分相关的 unigenes 有 18 类,其中与细胞组分、细胞器、细胞器组分、膜结构和复杂

大分子相关的 unigenes 所占比例分别是 82.91%, 53.38%, 45.25%, 30.62% 和 30.62%。而与分子功能相关的 unigenes 有 14 类,最多的是结合相关的,占 67.82%,分子功能中具有催化活性的 unigenes 占 48.14%。

将花椒窄吉丁的 unigenes 与 KOG 数据库进行比对并根据其功能进行分类统计 (图 6),注释到 24 614 个 unigenes 共 25 类。注释到 unigene 最多的功能类别是普通功能,共 4 666 个,其次为翻译后修饰、蛋白质折叠和分子伴侣功能共 2 548 个,而注释到最少的是与细胞移动相关,仅 40 个。所注释的功能涉及到花椒窄吉丁的绝大多数生命活动。

将 80 320 个 unigenes 映射到 KEGG 数据库中,注释为代谢通路相关的序列共有 21 104 条,占总数的 26.27%,把这些代谢通路分为五大类,分别是细胞过程、环境信息处理过程、遗传信息处理过程、新陈代谢过程和有机体系统过程;共涉及到 35 种代谢途径,其中注释的 unigene 最多的是翻译途径,共涉及到 2 037 条 unigenes,占 18.44%;其次为信号转导途径,共涉及到 1 910 条 unigenes,占 17.29%;第 3 为碳水化合物代谢,共涉及到 1 336 条 unigenes,占 12.09%。其他更多的代谢途径涉及运输和分解代谢、氨基酸代谢等 (图 7)。

对花椒窄吉丁转录组数据查找并对查找到的候选序列在 NR 数据库中进行比对验证后注释到 8 个气味受体 (odorant receptor, OR) 基因、7 个离子型受体 (ionotropic receptor, IR) 基因和 6 个味觉受体 (gustatory receptor, GR) 基因,共计 21 个 (表 5, 6)。根据基因命名规则分别对其命名为 AzanOR1 - 8, AzanIR1 - 7 和 AzanGR1 - 6。

2.4 花椒窄吉丁转录组化学感受相关基因的系统发育分析

图 8 所示,花椒窄吉丁的 OR 与白蜡窄吉丁 *Agrilus planipennis*, 黄粉虫 *Tenebrio molitor*, 暗黑鳃金龟 *Holotrichia parallela*, 大猿叶甲 *Colaphellus bowringi*, 大斑沟芫菁 *Hycleus phaleratus* 和星天牛 *Anoplophora chinensis* 6 种鞘翅目昆虫 OR 的系统发育树,从图可知 AzanOR1, AzanOR2 和 AzanOR6 与白蜡窄吉丁 OR 聚类在一起,相似度高, AzanOR3 与大斑沟芫菁 OR 相似度比较高,而 AzanOR4 和 AzanOR5 则与暗黑鳃金龟 OR 的相似度较高;图 9 是花椒窄吉丁 IR 与白蜡窄吉丁 *Agrilus planipennis*, 大猿叶甲 *Colaphellus bowringi*, 暗黑鳃金龟 *Holotrichia parallela*, 山松甲虫 *D. ponderosae*, 黄粉虫 *Tenebrio*

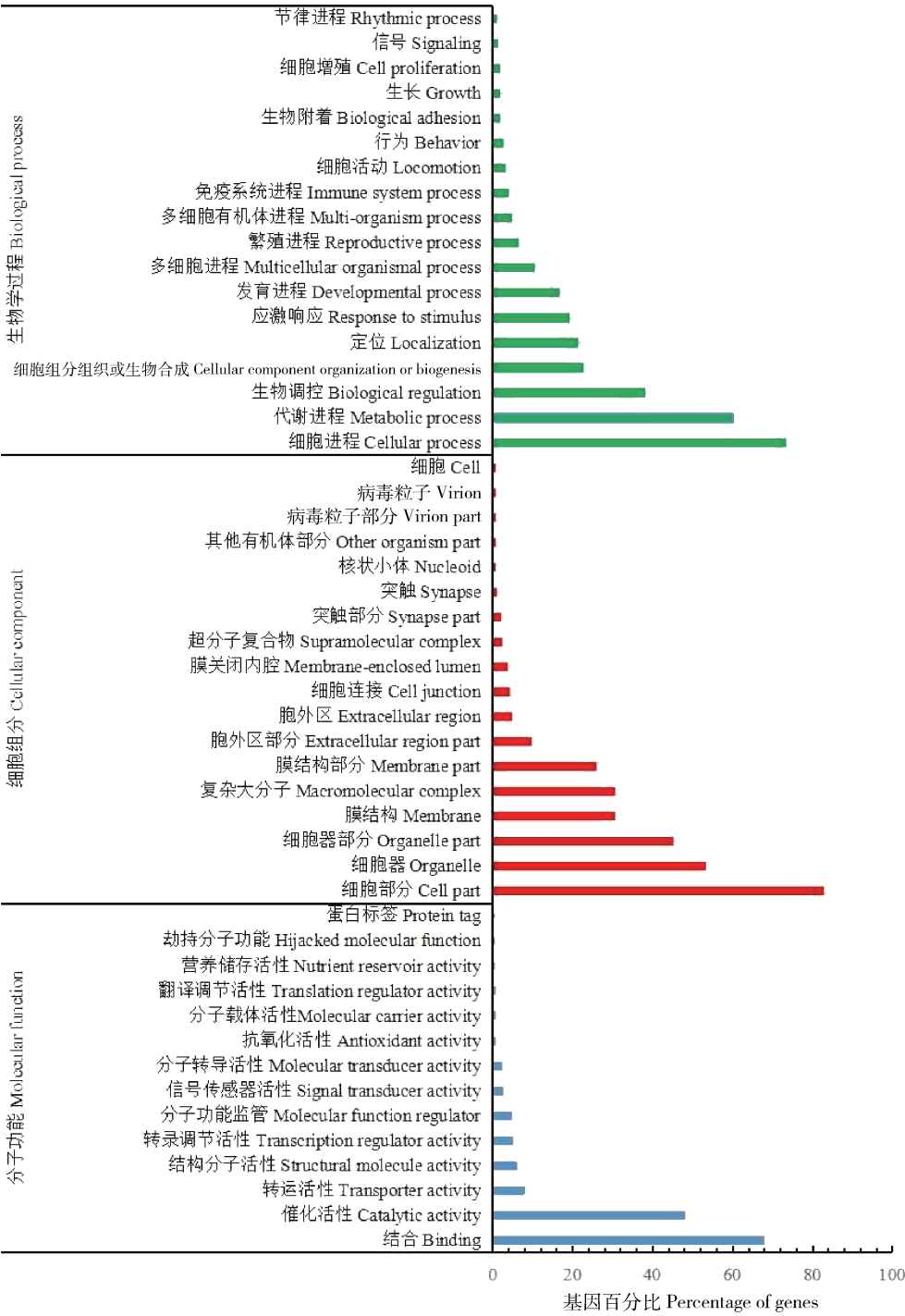


图 5 花椒窄吉丁 unigenes GO 分类

Fig. 5 GO classification of *Agrilus zanthoxylumi* unigenes

molitor 和大斑沟芫菁 *Hycleus phaleratus* IR 的系统发育树,从中可知花椒窄吉丁 IR 与白蜡窄吉丁、大斑沟芫菁、暗黑鳃金龟和大猿叶甲 IR 的相似度较高;图 10 是花椒窄吉丁的 GR 与白蜡窄吉丁 *Agrilus planipennis*, 大猿叶甲 *Colaphellus bowringi*, 大斑沟芫菁 *Hycleus phaleratus*, 星天牛 *Anoplophora chinensis*, 赤拟谷盗 *Tribolium castaneum* 和铜绿丽金龟

Anomala corpulenta GR 的系统发育树,花椒窄吉丁 GR 与白蜡窄吉丁、大猿叶甲和赤拟谷盗 GR 的相似度较高。这 3 类受体都与花椒窄吉丁的近源种之一的白蜡窄吉丁 GR 的相似度较高,在一定程度上说明了花椒窄吉丁转录组数据基因的准确性,也为今后这类基因的鉴定提供参照和思路。

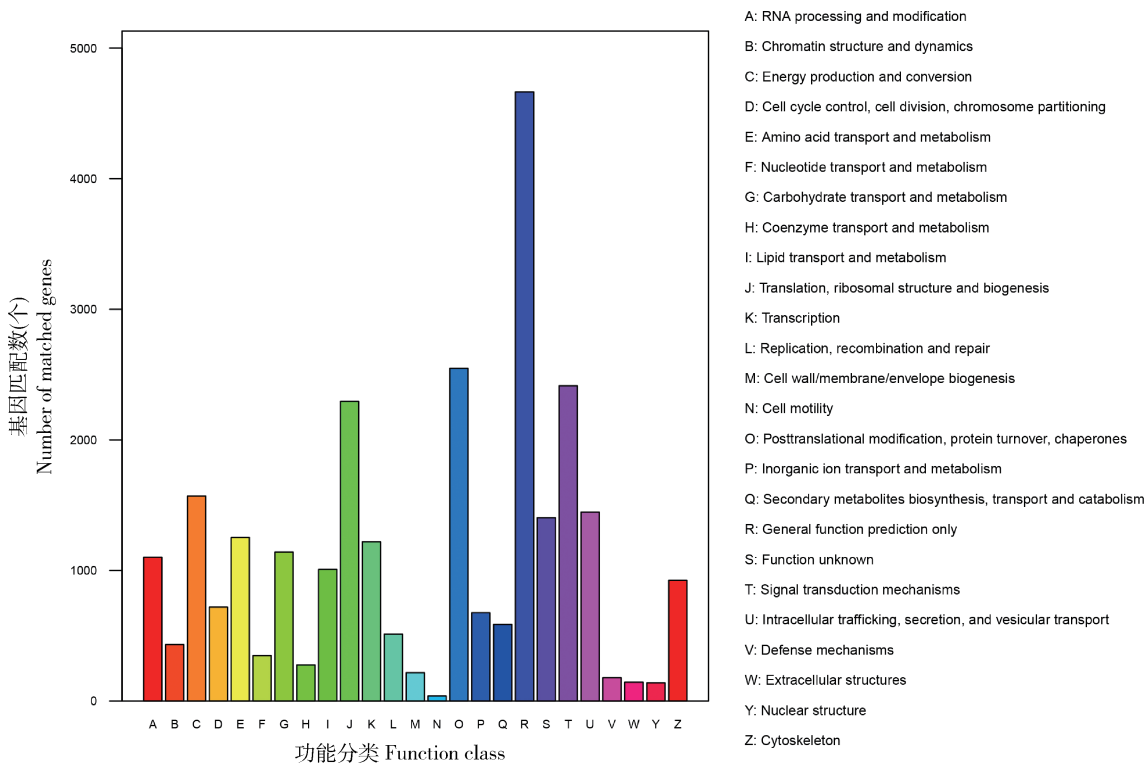


图 6 花椒窄吉丁 unigenes 的 KOG 功能分类
Fig. 6 KOG function classification of *Agrilus zanthoxylum* unigenes

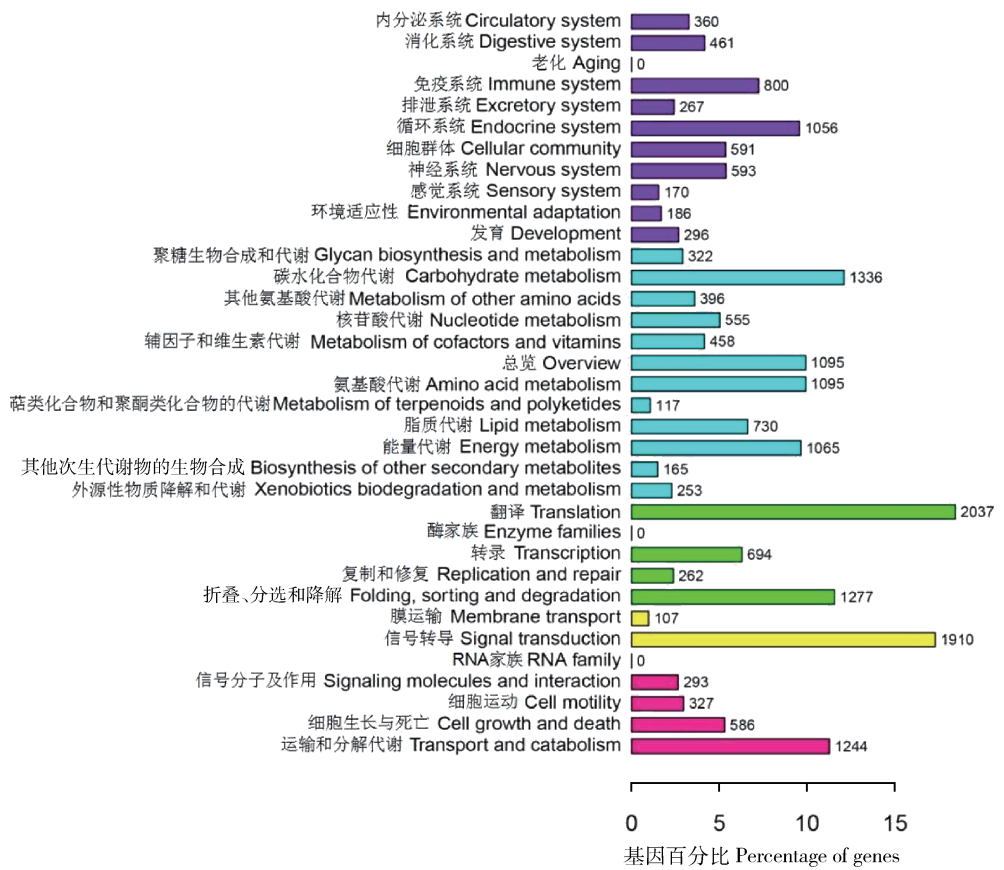


图 7 花椒窄吉丁 unigenes KEGG 代谢通路图
Fig. 7 KEGG metabolic pathway map of *Agrilus zanthoxylum* unigenes

表 5 花椒窄吉丁转录组化学感受相关基因注释结果统计

Table 5 Annotated chemoreception-related genes in the transcriptome of *Agrilus zanthoxylumi*

化学感受相关基因 Chemoreception-related genes	注释的基因数量 Number of annotated unigenes
气味受体基因 Odorant receptor (OR) gene	8
离子型受体基因 Ionotropic receptor (IR) gene	7
味觉受体基因 Gustatory receptor (GR) gene	6

表 6 花椒窄吉丁转录组化学感受相关基因 NR 中的注释结果

Table 6 Chemoreception-related genes in *Agrilus zanthoxylumi* transcriptome annotated in the NR database

基因 Genes	基因编号 Gene ID	NR 评分 NR score	NR 估算值 NR estimated value	描述 Description
<i>AzanOR1</i>	TRINITY_DN29115_c0_g1	671	E;0	PREDICTED:odorant receptor coreceptor [<i>Tribolium castaneum</i>]
<i>AzanOR2</i>	TRINITY_DN31824_c5_g1	111	E;4e-24	Odorant receptor OR11[<i>Colaphellus bowringi</i>]
<i>AzanOR3</i>	TRINITY_DN24144_c0_g1	313	E;3e-99	Odorant receptor 14 [<i>Anomala corpulenta</i>]
<i>AzanOR4</i>	TRINITY_DN30901_c0_g1	384	E;2e-118	Hypothetical protein D910_08613 [<i>Dendroctonus ponderosae</i>]
<i>AzanOR5</i>	TRINITY_DN33665_c3_g1	90.9	E;2e-16	PREDICTED: uncharacterized protein LOC103313324 [<i>Tribolium castaneum</i>]
<i>AzanOR6</i>	TRINITY_DN31824_c5_g1	111	E;4e-24	Odorant receptor OR11 [<i>Colaphellus bowringi</i>]
<i>AzanOR7</i>	TRINITY_DN25958_c1_g3	120	E;4e-28	Odorant receptor 14 [<i>Anomala corpulenta</i>]
<i>AzanOR8</i>	TRINITY_DN25958_c1_g2	105	E;8e-25	Odorant receptor 14 [<i>Anomala corpulenta</i>]
<i>AzanIR1</i>	TRINITY_DN29513_c1_g1	725	E;0	Ionotropic receptor 8a[<i>Colaphellus bowringi</i>]
<i>AzanIR2</i>	TRINITY_DN29688_c0_g1	1 360	E;0	Chemosensory ionotropic receptor IR6 [<i>Tenebrio molitor</i>]
<i>AzanIR3</i>	TRINITY_DN26923_c1_g4	166	E;3e-45	Chemosensory ionotropic receptor IR3, partial [<i>Tenebrio molitor</i>]
<i>AzanIR4</i>	TRINITY_DN28534_c3_g2	660	E;0	Chemosensory ionotropic receptor 21a, partial [<i>Anomala corpulenta</i>]
<i>AzanIR5</i>	TRINITY_DN27218_c0_g1	325	E;5e-96	Chemosensory ionotropic receptor IR5 [<i>Tenebrio molitor</i>]
<i>AzanIR6</i>	TRINITY_DN28534_c3_g2	660	E;0	Chemosensory ionotropic receptor 21a, partial [<i>Anomala corpulenta</i>]
<i>AzanIR7</i>	TRINITY_DN29688_c0_g1	1 360	E;0	Chemosensory ionotropic receptor IR6 [<i>Tenebrio molitor</i>]
<i>AzanGR1</i>	TRINITY_DN34348_c2_g2	66.2	E;6e-09	PREDICTED: gustatory receptor for sugar taste 43a-like [<i>Stomoxys calcitrans</i>]
<i>AzanGR2</i>	TRINITY_DN14380_c0_g1	135	E;6e-33	Gustatory receptor candidate 9[<i>Tribolium castaneum</i>]
<i>AzanGR3</i>	TRINITY_DN27483_c0_g3	221	E;1e-64	Gustatory receptor candidate 29[<i>Tribolium castaneum</i>]
<i>AzanGR4</i>	TRINITY_DN29925_c1_g3	157	E;2e-39	PREDICTED: putative gustatory receptor 28b [<i>Tribolium castaneum</i>]
<i>AzanGR5</i>	TRINITY_DN22765_c0_g1	210	E;2e-62	Hypothetical protein YQE_06884, partial [<i>Dendroctonus ponderosae</i>]
<i>AzanGR6</i>	TRINITY_DN27755_c0_g2	无 Null	无 Null	无 Null

2.5 花椒窄吉丁转录组化学感受相关基因的表达量分析

对 3 类化学感受相关基因在花椒窄吉丁成虫不同组织中(由于雌虫头部转录组数据缺失,因此头部数据在此不加以分析和讨论)的 RPKM 值进行对比发现,3 类化学感受相关基因在雌、雄成虫触角中表达量相对其他组织中要高,且 *AzanOR1* 表达量雌、雄差异最为显著;胸部中表达量雌、雄差异最显著的为 *AzanOR4*;腹部中 3 类化学感受相关基因表达量雌、雄差异最显著的为 *AzanOR5*;雌、雄翅中 3 类化学感受相关基因表达量差异最显著的为

AzanIR5 和 *AzanIR7*;而雌、雄成虫足中 3 类化学感受相关基因表达量没有差异(图 11)。

3 讨论

新一代高通量测序技术的不断发展,已彻底改变了转录组学的研究,使 RNA-Seq 无需预先设计探针即可对特定条件下任何生物生长发育阶段整体转录活动进行测序,准确探测到各种条件下的基因表达情况,发现了许多未知的分子调控机制(刘红亮等, 2013; 朱志明等, 2016; 王文龙等, 2017),一些

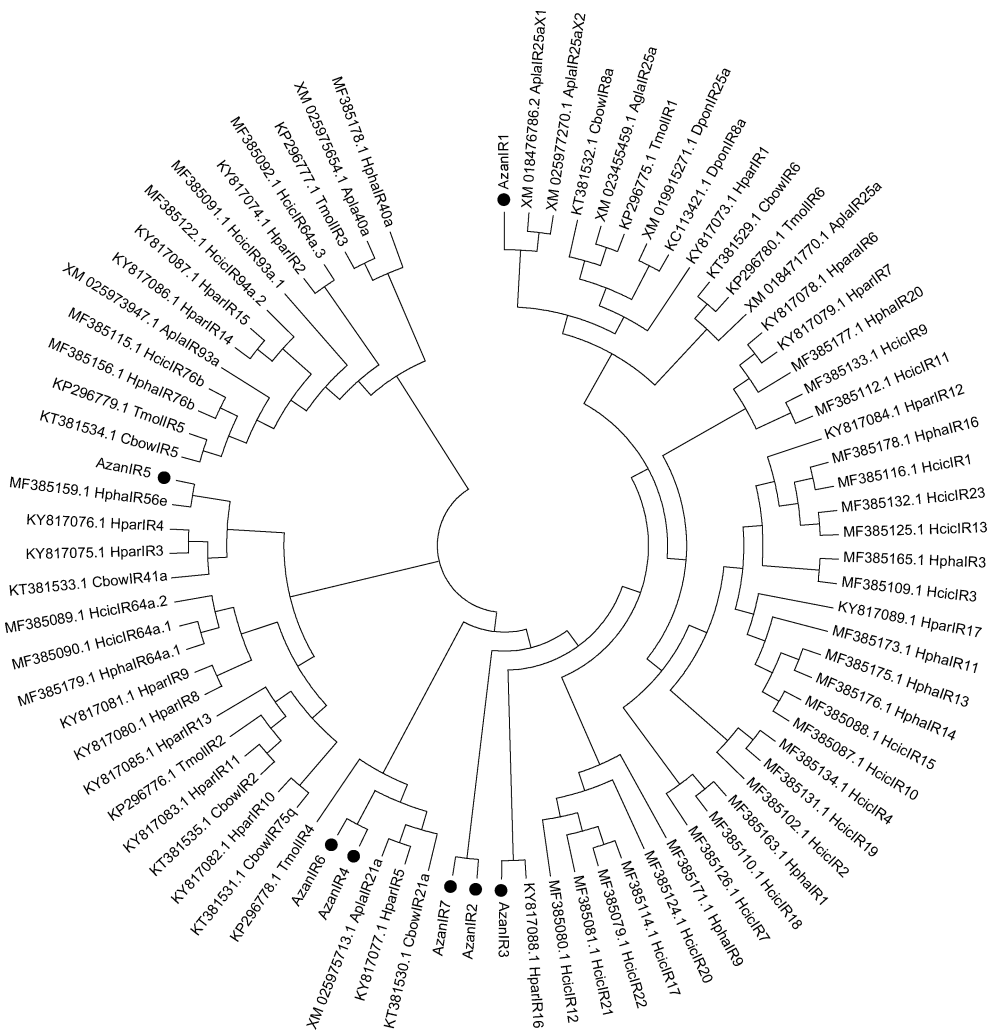


图9 邻接法构建的基于氨基酸序列的花椒窄吉丁和其他物种 IR 系统发育树

Fig. 9 Phylogenetic tree of IRs from *Agrilus zanthoxylumi* and other insect species by using neighbor-joining method based on the amino acid sequence

IR 来源物种 Source species of IRs: AplanIR: 白蜡窄吉丁 *Agrilus planipennis*; CbowIR: 大猿叶甲 *Colaphellus bowringi*; HparIR: 暗黑鳃金龟 *Holotrichia parallela*; DponIR: 山松甲虫 *Dendroctonus ponderosae*; TmolIR: 黄粉虫 *Tenebrio molitor*; HphaIR: 大斑沟壳菁 *Hyleus phaleratus*. 进化树中序列名为 GenBank 登录号 + 蛋白名。以bootstrap = 1 000构建进化树。The names of sequences in the phylogenetic tree are GenBank accession no. plus protein name. Bootstrap support values presented as percentages are based on 1 000 replicates.

数据库中,注释到代谢通路 35 个,其相关的序列共有 21 104 条,占总数的 26.27%,把这些代谢通路分为五大类,分别是细胞过程、环境信息处理过程、遗传信息处理过程、新陈代谢过程和有机体系统过程,这些代谢通路的分类与已报道的研究中并没有什么不同。本研究对花椒窄吉丁转录组原始数据的公开及分析,为种属鉴定及遗传多样性的分析提供了基础资料,同时也为了解其基因功能、生物学特征以及相关的生理学研究提供数据参考。

本研究首次从花椒窄吉丁转录组数据中挖掘到 8 个气味受体基因、7 个离子型受体基因和 6 个味觉受体基因,共计 21 个。味觉受体主要包含二氧化碳

检测受体(Kwon *et al.*, 2006),然而,大多数 GRs 在味觉器官的味觉受体神经元中表达并参与接触化学感受,这些 GRs 通常检测不同的糖、苦味化合物和接触信息素(Robertson and Kent, 2009)。昆虫 ORs 是含 7 个跨膜结构域的蛋白(Clyne *et al.*, 1999),与 GRs 是同一个受体基因超家族的(Vosshall *et al.*, 1999),但 ORs 和 GRs 相互之间的序列相关性较小(Robertson *et al.*, 2003)。而 IRs 是一类参与化学感受的典型受体(Benton *et al.*, 2009),它们主要在突触通讯中起作用。不同昆虫具有不同数目的气味接受相关基因,与花椒窄吉丁近源的苹果小吉丁 *A. mali* 中化学感受相关基因分别有 17 条 OR 基因、17

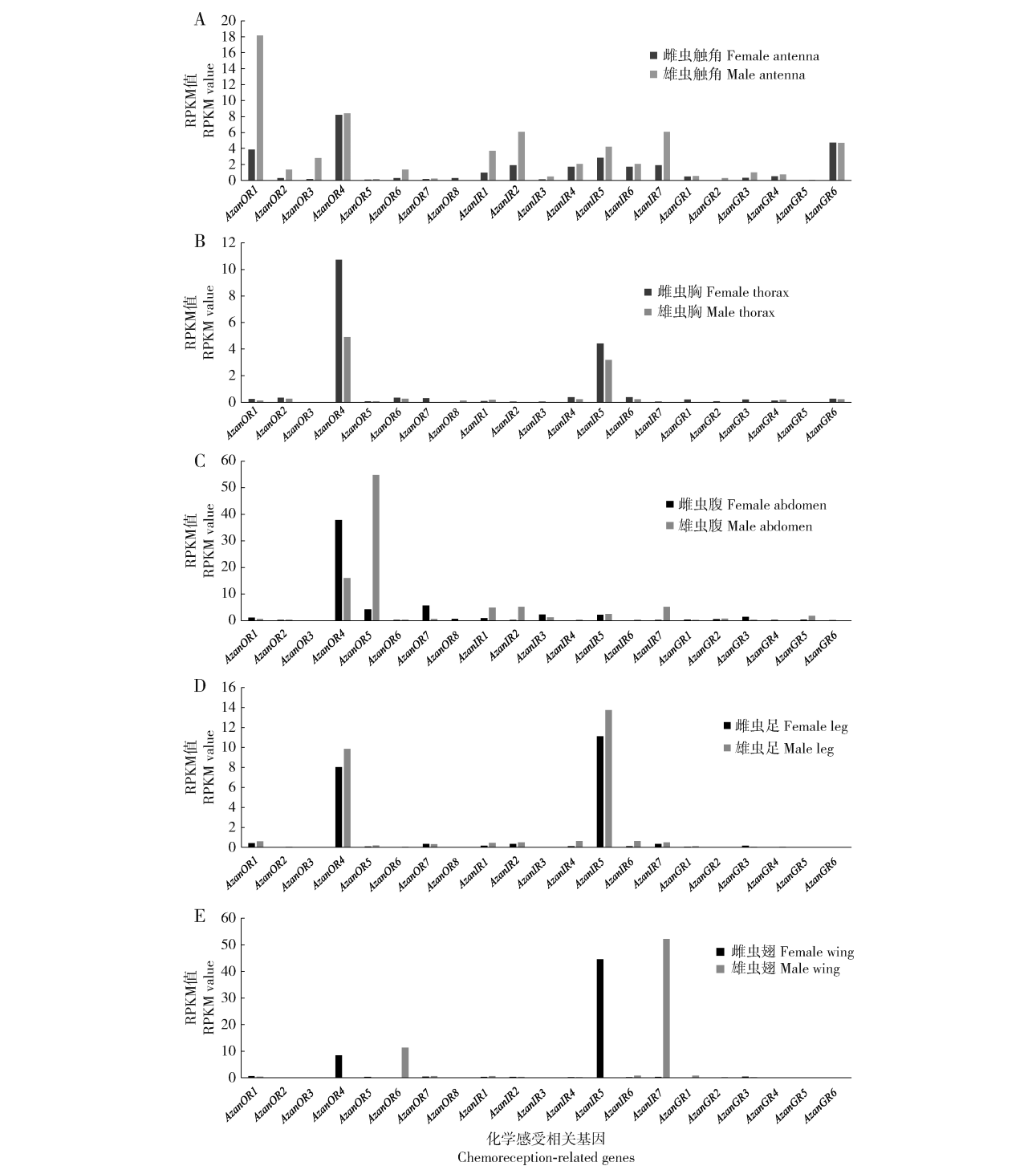


图 11 花椒窄吉丁化学感受相关基因在成虫不同组织中的 RPKM 值

Fig. 11 The RPKM values of chemoreception-related genes in different tissues of *Agrilus xanthoxylumi* adults

OR: 气味受体 Odorant receptor; IR: 离子型受体 Ionotropic receptor; GR: 味觉受体 Gustatory receptor. A: 触角 Antenna; B: 胸 Thorax; C: 腹 Abdomen; D: 足 Leg; E: 翅 Wing. 本研究采用 RPKM (每百万 reads 中来自某一基因每千碱基长度的 reads 数) 来表示基因的表达量,即图中纵坐标 RPKM 值表示基因的相对表达量。In this study, the number of reads per million from a certain gene per 1 000 base length (RPKM) was used to represent the gene expression level, that is, the ordinate RPKM value in the figure represents the relative expression level of the gene.

N, 2011. Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol. Ecol. Res.*, 11 (4): 650 – 661.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 9(7): 644 – 652.

Hansen KD, Irizarry RA, Wu ZJ, 2012. Removing technical variability

- in RNA-seq data using conditional quantile normalization. *Biostatistics*, 13(2): 204–216.
- He HL, Bin SY, Wu ZZ, Lin JT, 2012. Transcriptome characteristics of *Phyllotreta striolata* (Fabricius) (Coleoptera: Chrysomelidae) analyzed by using Illumina's Solexa sequencing technology. *Acta Entomol. Sin.*, 55(1): 1–11. [贺华良, 宾淑英, 吴仲真, 林进添, 2012. 基于 Solexa 高通量测序的黄曲条跳甲转录组学研究. 昆虫学报, 55(1): 1–11]
- He K, 1991. Wound fluid glue and treatment of *Zanthoxylum bungeanum*. *Gansu Agric. Sci. Technol.*, (1): 31. [何凯, 1991. 花椒树的伤口流胶及治疗. 甘肃农业科技, (1): 31]
- Kang KG, Wang YP, Qiang L, Shi HQ, Chen H, 2007. A study on chemical control test of *Agrilus zanthoxylumi* Hou and Feng. *J. Northwest For. Univ.*, 22(6): 105–107. [康克功, 王永平, 强磊, 石和芹, 陈辉, 2007. 花椒窄吉丁化学防治研究. 西北林学院学报, 22(6): 105–107]
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR, 2007. The molecular basis of CO₂reception in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 104(9): 3574–3578.
- Liu HL, Zheng LM, Liu QQ, Quan FS, Zhang Y, 2013. Studies on the transcriptomes of non-model organisms. *Hereditas (Beijing)*, 35(8): 955–970. [刘红亮, 郑丽明, 刘青青, 权富生, 张涌, 2013. 非模式生物转录组研究. 遗传, 35(8): 955–970]
- Nguyen CT, Lim S, Lee JG, Lee EJ, 2017. VcBBX, VcMYB21 and VcR2R3MYB transcription factors are involved in UV-B-induced anthocyanin biosynthesis in the peel of harvested blueberry fruit. *J. Agric. Food Chem.*, 65(10): 2066–2073.
- Robertson HM, Kent LB, 2009. Evolution of the gene lineage encoding the carbon dioxide receptor in insects. *J. Insect Sci.*, 9: 19.
- Robertson HM, Warr CG, Carlson JR, 2003. Molecular evolution of the insect chemoreceptor gene super family in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, 100(s2): 14537–14542.
- Sangwan RS, Tripathi S, Singh J, Namoliya LK, Sangwan NS, 2013. *De novo* sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene*, 525(2): 58–76.
- Sun YZ, Li ZF, 1988. A preliminary study of *Agrilus zanthoxylumi*. *Plant Prot.*, 14(2): 23–24. [孙益知, 李忠锋, 1988. 花椒吉丁虫的初步研究. 植物保护, 14(2): 23–24]
- Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R, 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*, 96: 725–736.
- Wang WL, Feng CC, Hong M, Yue JW, Huhebateer, Liu CX, 2017. The comparative transcriptome analysis of *Parabronema skrjabini* at different developmental stages. *Sci. Agric. Sin.*, 50(23): 4644–4655. [王文龙, 冯陈晨, 红梅, 岳建伟, 呼和巴特儿, 刘春霞, 2017. 不同发育阶段斯氏副柔线虫比较转录组学分析. 中国农业科学, 50(23): 4644–4655]
- Wang XC, Tan HL, Chen Z, Meng LZ, Wang WB, Fan SC, 2015. Assembly and characterization of the transcriptome and development of SSR markers in *Forsythia suspensa* based on RNA-Seq technology. *Sci. Sin. Vit.*, 45(3): 301–310. [王兴春, 谭河林, 陈钊, 孟令芝, 王文斌, 范圣此, 2015. 基于 RNA-Seq 技术的连翘转录组组装与分析及 SSR 分子标记的开发. 中国科学: 生命科学, 45(3): 301–310]
- Wu H, 2006. Biological characteristics and control of *Agrilus zanthoxylum*. *Chin. Bull. Entomol.*, 43(2): 236–239. [吴海, 2006. 花椒窄吉丁的生物学特性及防治. 昆虫知识, 43(2): 236–239]
- Xu L, Yang P, Yuan S, Feng Y, Xu H, Cao Y, Ming J, 2016. Transcriptome analysis identifies key candidate genes mediating purple ovary coloration in Asiatic hybrid lilies. *Int. J. Mol. Sci.*, 17(11): 1881–1888.
- Yang F, Huang LH, Zhang AB, 2014. High-throughput transcriptome sequencing technology and its applications in Lepidoptera. *Acta Entomol. Sin.*, 57(8): 991–1000. [杨帆, 黄立华, 张爱兵, 2014. 高通量转录组测序技术及其在鳞翅目昆虫上的应用. 昆虫学报, 57(8): 991–1000]
- Yang F, Shi ZY, Bai SL, Zhang AB, 2014. Comparative studies on species identification of Noctuoidea moths in two nature reserve conservation zones (Beijing, China) using DNA barcodes and thin-film biosensor chips. *Mol. Ecol. Resour.*, 14(1): 50–59.
- Yuan LF, 2016. The Preliminary Study on Mechanism of Odor Perception in *Agrilus zanthoxylumi* to Volatiles from *Zanthoxylum bungeanum*. MSc Thesis, Northwest A&F University, Yangling, Shaanxi. [袁丽芳, 2016. 花椒窄吉丁对寄主挥发物化学感受机制的初步研究. 陕西杨凌: 西北农林科技大学硕士学位论文]
- Zhang MC, Yin J, Cao YZ, Wu JX, Li KB, 2014. Transcriptome sequencing and its application in insect research. In: Green Ecological Sustainable Development and Plant Protection – Proceedings of the 12th National Congress and Annual Conference of the Chinese Society for Plant Protection. China Society of Plant Protection. 120–127. [张美翠, 尹姣, 曹雅忠, 仝均祥, 李克斌, 2014. 转录组测序及其在昆虫研究中的应用. 见: 绿色生态可持续发展与植物保护——中国植物保护学会第十二次全国会员代表大会暨学术年会论文集. 中国植物保护学会学术年会. 120–127]
- Zhang W, 2009. Research on Natural Enemies of *Agrilus zanthoxylumi*. MSc Thesis, Northwest A&F University, Yangling, Shaanxi. [张伟, 2009. 花椒窄吉丁天敌研究. 陕西杨凌: 西北农林科技大学硕士学位论文]
- Zhou H, Zhang X, Liu TY, Yu FX, 2012. Data analysis and gene discovery for high throughput transcriptome sequencing. *Jiangxi Sci.*, 30(5): 607–611. [周华, 张新, 刘腾云, 余发新, 2012. 高通量转录组测序的数据分析与基因发掘. 江西科学, 30(5): 607–611]
- Zhu ZM, Chen HP, Lin RL, Miao ZW, Xin QW, Li L, Zhang DQ, Zheng NZ, 2016. Transcriptome analysis of ovary tissue in early laying period and egg laying peak period of Shanma ducks. *Sci. Agric. Sin.*, 49(5): 998–1007. [朱志明, 陈红萍, 林如龙, 缪中纬, 辛清武, 李丽, 张丹青, 郑嫩珠, 2016. 山麻鸭开产期和产蛋高峰期卵巢组织转录组分析. 中国农业科学, 49(5): 998–1007]